Kinetics and Mechanisms of Nucleophilic Displacements with Heterocycles as Leaving Groups. Part 19.[1] Chemometric Investigation of the Simultaneous Dependence of S_N2 Rates on Alkyl Group Structure and Leaving Group Nucleofugacity

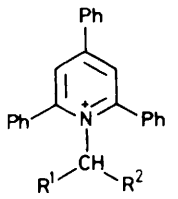
Giuseppe Musumarra* and Mario Bruno
Istituto Dipartimentale di Chimica e Chimica Industriale, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy
Alan R. Katritzky* and Kumars Sakizadeh
Department of Chemistry, University of Florida, Gainesville, Florida 32611, U.S.A.
Sergio Alunni and Sergio Clementi*
Dipartimento di Chimica, Università di Perugia, Via Elce di Sotto 10, 06100 Perugia, Italy

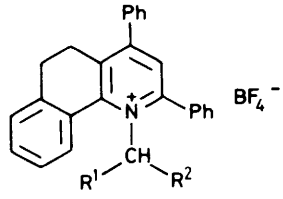
The preparation and kinetics of nucleophilic displacement are reported for four β-branched primary alkyl groups attached to neutral nucleofuges. Principal-component analysis on a set of 10 nucleophilic substitution reactions with neutral and anionic nucleofuges finds that the first principal component accounts for 70% of the variance and confirms that the tri- and penta-cyclic nucleofuges are similar to chloride ion in leaving-group activity. Partial least-squares analysis shows that the nucleophilic displacement rates for the tricyclic derivatives (**2**) depends on substituent shape (as measured by the Verloop parameter) rather than on size as measured by E_s. The σ* and polarizability terms are also important.

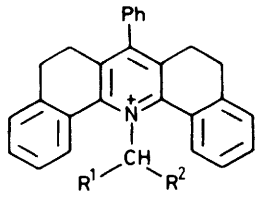
Previous studies on the kinetics and the mechanism of the reactions of *N*-benzyl, *N*-allyl, *N*-n-alkyl, and *N*-s-alkyl derivatives in series (**1**)—(**3**) with piperidine in chlorobenzene[2,3] have shown that although the S_N2 rates invariably increase in the order (**1**) < (**2**) < (**3**), for different nitrogen substituents, the rate enhancement varies considerably. This enhancement grows on increasing the size of the nitrogen substituent and the steric shape of the nitrogen substituent is also important. However, for each series, the same rate sequence was found: benzyl > methyl ≃ s-alkyls > continuous chain primary alkyls ≃ neopentyl. This sequence contrasts with that generally accepted for S_N2 rates;[4,5] benzyl > methyl > primary alkyl > secondary alkyl ≫ neopentyl. Only compounds with secondary *N*-substituents in series (**1**)—(**3**) exhibited a significant first-order component. No branched primary alkyl groups were studied other than neopentyl.

Following these studies we now report the displacement kinetics for some analogous compounds of series (**2**) and (**3**) with branched primary *N*-alkyl and *N*-(cycloalkylmethyl) substituents. These new results, together with previous kinetic data on the S_N2 rate dependence on both alkyl and the nucleofuge in our nucleophilic displacements with nitrogen heterocycles as leaving groups and in other conventional bimolecular substitutions available from the literature, provide a suitable data set for multivariate statistical analysis. We therefore now also report a chemometric investigation of the above rate data matrix, using principal-component analysis (PCA) and the recently developed method of partial least-squares (PLS) analysis, with the aim of studying the simultaneous dependence of S_N2 rates on alkyl-group structure and leaving-group nucleofugacity.

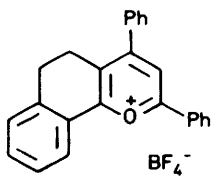
Preparation of Compounds.—5,6-Dihydro-2,4-diphenylbenzo[*h*]chromylium tetrafluoroborate (**4**) and 5,6,8,9-tetrahydro-7-phenyldibenzo[*c,h*]xanthylium tetrafluoroborate (**5**)[6] were reacted[6,7] with isobutylamine, cyclohexylmethylamine, 2-methylbutylamine, and cyclopropylmethylamine to give the corresponding pyridinium tetrafluoroborates (**2m—p**) and (**3m—p**) (Table 1). The amines react faster (3—5 h) with


(1)


(2)


(3)

	R^1	R^2
a;	H	Ph
b;	H	$CH{=}CH_2$
c;	H	H
d;	H	Me
e;	H	Et
f;	H	Pr
g;	H	Bu
h;	H	n-C_5H_{11}
i;	H	n-C_6H_{13}
j;	H	Bu^t
k;	Me	Me
l;	Me	Et
m;	H	Pr^i
n;	H	cyclo-C_3H_5
o;	H	Bu^s
p;	H	cyclo-C_6H_{11}
q;	H	$CH{=}CHCH_3$


(4)

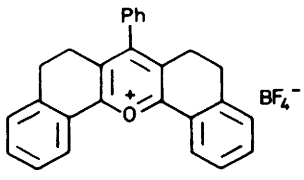

(5)

Table 1. Preparation of pyridinium tetrafluoroborates from pyryliums

Compd.	*N*-Substituent	Reaction time (h)	Recryst. solvent	M.p. (°C)	Yield (%)	Found (%) C	H	N	Formula	Required (%) C	H	N
(2m)	CH_2CHMe_2	6[a]	Me_2CO–Et_2O	224	76	72.8	5.8	2.9	$C_{29}H_{23}BF_4N$	72.9	5.8	2.9
(2n)	CH_2-cyclopropyl[b]	16	Me_2CO–Et_2O	133	92	73.1	5.6	2.8	$C_{29}H_{26}BF_4N$	73.2	5.5	2.9
(2o)	$CH_2CHMeEt$	18	n-C_6H_{12}	155	76	73.1	6.1	2.7	$C_{30}H_{30}BF_4N$	73.3	6.1	2.8
(2p)	CH_2-cyclohexyl	18	Me_2CO–Et_2O	203	78	74.1	6.2	2.7	$C_{32}H_{32}BF_4N$	74.2	6.2	2.7
(3m)	CH_2CHMe_2	4[a]	Me_2CO–Et_2O	236	84	73.7	6.1	2.7	$C_{31}H_{30}BF_4N$	73.9	5.9	2.8
(3n)	CH_2-cyclopropyl[b]	5	Me_2CO–Et_2O	106	93	74.2	5.5	2.7	$C_{31}H_{28}BF_4N$	74.2	5.6	2.8
(3o)	$CH_2CHMeEt$	3	Me_2CO–Et_2O	207	89		*c*		$C_{32}H_{32}BF_4N$	74.3	6.2	2.7
(3p)	CH_2-cyclohexyl	3—4	Me_2CO–Et_2O	236	92		*c*		$C_{34}H_{34}BF_4N$	75.1	6.2	2.6

[a] AcOH (1 mmol) was added to the reaction mixture after 2 h. [b] Hydrochloride salt of this amine was used which was dissolved in CH_2Cl_2 (2 ml) with the addition of EtOH (0.1 ml). [c] Also characterized by ^1H and ^{13}C n.m.r. spectroscopy.

Table 2. ^1H N.m.r. spectral data[a,b]

Compound	Aromatic: Py-ring δ	H	Multiplet δ	H	N-C*H*$_2$ 2 H, d δ	*J* (Hz)	CH_2-CH_2 δ	H	NCH_2CH 1 H, m δ	CH_2 m δ	H	CH_3 δ	H	m	*J*
(2m)	8.10	1	7.5	13	4.90	7	2.85	4	1.53			0.40	6	d	7
(2n)	8.00	1	7.60	13	5.00	7	2.98	4	0.75	0.45	2				
(2o)	8.10	1	7.68	13	5.01	7	2.95	4	1.30	0.80	2	0.47	6	m	
(2p)	8.00	1	7.60	13	4.90	6	2.90	4	1.40	0.90	10				
(3m)	8.25	2	7.55	11	5.16	7	2.80	8	1.60			0.45	6	d	7
(3n)	8.05	2	7.60	11	5.20	6	2.83	8	0.80	0.40—0.15	4				
(3o)	8.20	2	7.50	11	5.25	7	2.85	8	1.60—1.35	1.25—1.00	2	0.98—0.60	6	m	
(3p)	8.20	2	7.65	11	5.30	7	2.90	8	1.55	0.90	10				

[a] All the spectra were taken in $CDCl_3$ with the addition of 2—3 drops of CF_3CO_2H. [b] d = doublet, m = multiplet.

Table 3. Pseudo-first-order rate constants (10^5 k_{obs}/s^{-1}) for the reactions of compounds in series (**2**) and (**3**) with piperidine in chlorobenzene at 100 °C[a]

Compound	*N*-Substituent	Kinetic λ (nm)	[pip]/mol l^{-1} 0.08	0.16	0.24	0.32
(2m)	CH_2CHMe_2	354	0.465	0.860	1.37	1.80
(2n)	CH_2-cyclopropyl	350	53.4	77.5	101	126
(2o)	$CH_2CHMeEt$	354	0.400	0.782	1.12	1.51
(2p)	CH_2-cyclohexyl	354	0.290	0.717	1.04	1.39
(3m)	CH_2CHMe_2	392	4.21	10.5		21.5
(3n)	CH_2-cyclopropyl[b]	392		86.2	115	142
(3o)	$CH_2CHMeEt$	394	2.60	5.90	8.61	12.45
(3p)	CH_2-cyclohexyl	392	2.67		7.60	10.05

[a] [Substrate] = 6.4×10^{-5} (mol l^{-1}). [b] Additional kinetic runs: [piperidine] (10^5 k_{obs}/s^{-1}), 0.04 (42.6), 0.008 (28.7).

the pentacyclic pyrylium (**5**) than with the tricyclic system (**4**).

Compounds were characterized by elemental analysis (Table 1) and ^1H n.m.r. spectra. Pyridinium ring hydrogens, ethylene bridges, N*CH*$_2$, NCH_2CH, and other aliphatic protons appeared in the expected regions[6,8] with correct integrations (see Table 2).

Kinetic Results.—The reactions of cations (**2**) and (**3**) with an excess of piperidine (under pseudo-first-order conditions) were followed spectrophotometrically at 100 °C by measuring the disappearance of the cation.[2,9,10] Pseudo-first-order rate constants (k_{obs}) are recorded in Table 3 together with the kinetic wavelengths. Plots of k_{obs} against the nucleophile concentration gave straight lines; the slopes are considered to vary as k_2 and the intercepts as k_1, the second- and first-order rate constants for S_N2 and S_N1 nucleophilic substitutions respectively (see discussion in ref. 10). First- and second-order rate constants for the reactions of the tricyclic derivatives (**2m**—**p**) and of their pentacyclic analogues (**3m**—**p**) with piperidine in chlorobenzene at 100 °C are reported in Table 4.

No significant first-order component occurs for branched primary alkyl derivatives; however, a significant first-order rate is observed for the *N*-(cyclopropylmethyl) compounds (**2n**) and (**3n**).

Table 5 gives the logarithms of second-order rates relative to the ethyl compound for each leaving group. Comparative data are not available for cyclo-C_3H_5, Bu^s, cyclo-C_6H_{11} and $CH{=}CHCH_3$. However, the rate sequence for n-butyl, iso-butyl, and cyclohexylmethylene in both series (**2**) and (**3**) are in good agreement with those for the reactions of alkyl bromides with chloride ion in acetone–water[11] and with methoxide in methanol.[12] Second-order rate constants for cyclopropylmethyl compounds (**2n**) and (**3n**), however, are higher than those of the analogous cyclohexylmethyl derivatives (**2p**) and (**3p**), in contrast with the usual rate enhancement on increasing the ring size from three to six observed for the S_N2 reactions of bromo-

Table 4. First- (k_1) and second-order (k_2) rate constants for the reactions of compounds in series (**2**) and (**3**) with piperidine in chlorobenzene at 100 °C

Compound	*N*-Substituent	r^a	N^b	$10^3\ k_2$/l mol^{-1} s^{-1} [c]	Error (%)	$10^5\ k_1$/s^{-1} [c]	Error (%)	$\frac{1\,000\ k_1}{k_2 + 10k_1}$ [d]
(**2m**)	CH_2CHMe_2	0.999	4	0.0564 ± 0.0054	10	(−0.01 ± 0.12)		<16
(**2n**)	CH_2-cyclopropyl	0.9999	4	3.01 ± 0.08	3	29.1 ± 1.6	6	49
(**2o**)	$CH_2CHMeEt$	0.9996	4	0.0458 ± 0.0025	5	(0.04 ± 0.06)		<18
(**2p**)	CH_2-cyclohexyl	0.998	4	0.0453 ± 0.0056	12	(−0.05 ± 0.12)		<13
(**3m**)	CH_2CHMe_2	0.999	3	0.72 ± 0.12	18	(−1.3 ± 2.6)		<15
(**3n**)	CH_2-cyclopropyl	0.9996	5	3.62 ± 0.12	3	27.2 ± 2.4	9	43
(**3o**)	$CH_2CHMeEt$	0.998	4	0.40 ± 0.054	13	(−1 ± 1)		<33
(**3p**)	CH_2-cyclohexyl	0.9999	3	0.307 ± 0.003	1	0.211 ± 0.070	33	6

[a] Correlation coefficient. [b] Number of runs. [c] 90% Confidence limits. [d] % S_N1 reaction at [piperidine] 0.1M.

methylcycloalkanes with methoxide[12] and thiophenoxide[13] ions. Moreover, in the case of the cyclopropylmethyl derivatives, reaction also occurs by the S_N1 mechanism, the first-order rate constants being higher than those of the cyclohexylmethyl analogues.

In agreement with previous findings,[2,10] the S_N2 reaction of each pentacyclic derivative (**3**) is faster than that of the tricyclic analogue (**2**) with the rate enhancement varying widely for different nitrogen substituents. Thus the benzyl,[2] the n-pentyl,[2] the isobutyl, the 2-methylbutyl, and the cyclohexylmethyl compounds respond much more (14, 14, 13, 9, and 7 times, respectively) than the allyl[2] and the cyclopropylmethyl compounds (only 1.2 times) to the second annulation.

Multivariate Statistical Methods.—Chemometrics, the application of mathematical and statistical methods to chemistry,[14] is a new discipline which has developed parallel to the improvement of computing facilities over the last decade. Multivariate statistics is of proven utility in handling complex chemistry-related data sets.[15,16]

A data set suitable for a multivariate analysis consists of a table (matrix) where a number (M) of experimental values (variables) is collected for each of the N chemical compounds (objects). The geometrical interpretation of each object is a point in the M-dimensional space, where each variable defines an orthogonal axis. Accordingly, the data set has the form of N points in an M space. Multivariate methods seek for the structure of the data, *i.e.* they are aimed at recognising systematic patterns, if present. This research area is also called 'pattern recognition'.[14,15]

Pattern-recognition methods apply similarity criteria. Some of them are based on the Euclidean distance: the closer two points are in the M-space the more similar are the two objects. Other methods use as similarity criterion the fit to a unique mathematical model and are based on least-squares procedures. Among these, multiple regression analysis (MRA), principal-components analysis (PCA), and partial-least squares analysis (PLS) are particularly appropriate in physical organic chemistry, as they allow the description of the data by mathematical equations.

MRA[17] describes one selected dependent variable y_i as a function of a number of independent variables x_{ia} [equation (1)]. MRA is still the most popular multivariate approach, but

$$y_i = c_o = \sum_{a=1}^{A} c_a x_{ia} + e_i \qquad (1)$$

its use involves the following implicit assumptions. (1) All the variables x_{ia} are independent and error free (otherwise multicollinearity can give meaningless regression coefficients). (2) All the independent variables used are relevant to the problem. (3) There is an absence of non-random groupings of the data points (subgroups cannot be recognised).

In PCA no cause–effect relationship is assumed and rather than select one unique y_i all the M variables are treated in the same way.[18] The method seeks systematic variations in the data matrix to elucidate the structure of the objects in the M-space. No assumption is required about the variables, and the correlations between them determine the mathematical solution which consists of the simultaneous explanation of all objects by the variables. PCA selects the best model, with the minimum number of dimensions, to explain the data structure. The plots of the components against each other (also called eigenvector plots) illustrate the data structure and can be regarded as windows opened on the multidimensional data set.

The SIMCA method, a computer package developed at the University of Umea,[15,19–21] applies disjoint PC models to each class of homogeneous objects. The data matrix contains elements x_{ik}, where index i is used for the experimental measurements (variables) and index k for the chemical compounds (objects). Each element is described by equation (2), where the number A of significant cross-terms (components), and the parameters b_{ia}, t_{ak} are calculated by minimizing the squared residuals e_{ik}, after subtracting $\bar{x}_i$ (the mean value of the k experimental quantities x_i).

$$x_{ik} = \bar{x}_i + \sum_{a=1}^{A} b_{ib} t_{ak} + e_{ik} \qquad (2)$$

In this model, parameters $\bar{x}_i$ and b_{ia} depend only on the variables, and t_{ak} only on the compounds. The deviations from the model are expressed by the residuals e_{ik}. By scaling to the same variance (fixed to unity), the variables are all given the same importance in the analysis. The PCA then proceeds by model expansions to find the correct dimensionality A using the cross-validation technique.[22]

The relevance of each variable in describing the mathematical model is given by its modelling power ψ_i [equation (3)

$$\psi_i = 1 - s_i(A = A)/s_i(A = 0) \qquad (3)$$

where s_i is the residual standard deviation for each variable after A dimensions and after dimension zero].

When the normalization of raw data is done by autoscaling, the ψ_i values are strictly related to the b_i values for the first component. However, the modelling powers are more easy to interpret, since the b_i parameters are calculated under the constraint $\Sigma b_i^2 = 1$, which makes them very similar to each other. Nevertheless the b_i values provide the relative signs of the variables.

The SIMCA method has already been applied successfully in physical organic chemistry providing new insights on linear free-

Table 5. Logarithms of second-order relative rates for the reaction of *N*-alkyl- and *N*-benzylpyridiniums (**1**)—(**3**) with piperidine in chlorobenzene at 100 °C and for the reaction of benzyl and alkyl halides with nucleophiles

		Reaction number									
	Alkyl Group	1 Series (**1**)[a] + piperidine	2 Series (**2**)[a] + piperidine	3 Series (**3**)[a] + piperidine	4 RI + PhO^- [b]	5 RBr + EtO^- [c]	6 RI + NEt_3 [d]	7 RBr + Cl^- [e]	8 RBr + Br^- [e,f]	9 RCl + KI[g]	10 RBr + MeO^- [h]
a	Benzyl	2.205	2.389	2.822						1.903	
b	Allyl	1.505	1.450	0.924						1.505	
c	Me	1.301	0.580	0.568	0.653	1.255	1.041	0.568	1.881	1.903	0.954
d	Et	0	0	0	0	0	0	0	0	0	0
e	Pr^n		−0.432	−0.387	−0.409	−0.509	−0.721	−0.161	−0.187	−0.367	−0.432
f	Bu^n		−0.301	−0.444	−0.444	−0.638	−0.854			−0.398	−0.432
g	n-Pentyl		−0.959	−0.523	−0.796	−0.699				−0.276	
h	n-Hexyl		−1.150	−0.301	−0.469					−0.284	
i	n-Heptyl		−1.100	−0.432	−0.481		−1.00			−0.301	
j	Neopentyl		−1.000			−2.377		−5.221	−5.00		
k	Pr^i	0.663	0.322		−0.469		−1.699	−1.699	−2.00	−2.097	
l	Bu^s	0.556	1.114		−0.420					−2.097	
m	Bu^i		−1.367[i]	−0.969[i]	−0.854			−1.523			−1.131
n	CH_2-cyclopropyl		0.358[i]	−0.267[i]							
o	2-Methylbutyl		−1.460[i]	−1.224[i]							
p	CH_2-cyclohexyl		−1.464[i]	−1.339[i]							
q	γ-Methylallyl		2.025[j]								

[a] From ref. 2, except where otherwise stated. [b] In dry EtOH at 42.5 °C; from: D. Segaller, *J. Chem. Soc.*, 1913, **103**, 1154; 1914, **105**, 106. [c] In dry EtOH at 55 °C; from: C. K. Ingold, 'Structure and Mechanism in Organic Chemistry,' Bell, London, 1969, 2nd edn., pp. 432—436; I. Dostrovsky and E. D. Hughes, *J. Chem. Soc.*, 1946, 157; M. L. Dhar, E. D. Hughes, C. K. Ingold, and S. Masterman, *ibid.*, 1948, 2055. [d] In acetone at 100 °C, from refs. in footnote *c* and N. Menschutkin, *Z. Phys. Chem.*, 1890, **5**, 589. [e] In DMF at 25 °C, from: S. Hartshorn, 'Aliphatic Nucleophilic Substitution,' Cambridge University Press, Cambridge, 1973, p. 32. [f] In acetone at 25 °C, from: refs. in footnote *c* and P. B. D. de la Mare, *J. Chem. Soc.*, 1955, 3180. [g] In acetone at 50 °C, from: J. Hine, 'Physical Organic Chemistry,' McGraw Hill, New York, 1962, 2nd edn., p. 176; J. B. Conant and R. E. Hussey, *J. Am. Chem. Soc.*, 1925, **47**, 476; J. B. Conant, W. R. Kirner, and R. E. Hussey, *ibid.*, p. 488. [h] In MeOH at 80 °C, from: K. Okamoto, I. Nitta, T. Imoto, and H. Shingh, *Bull. Chem. Soc. Jpn.*, 1967, **40**, 1905. [i] This work. [j] Extrapolated value, from ref. 3.

energy relationships.[23] Other applications included multivariate analysis of solvolysis rate data and the substituent 'descriptors',[24] ^{13}C n.m.r. studies,[25–29] investigations on the solvent effects,[30,31] and on the relationships between chemical structure and biological activities.[32]

PCA is superior to MRA whenever uncertainty exists regarding which variables significantly affect the problems; however, it is not aimed at finding out cause–effect relationships. A correct statistical approach aimed at this objective, able to cope both with the interpretation of results and the prediction of unmeasured data, is provided by the recently developed method called partial least-squares (PLS) analysis.[20,21,33,34] A dependent variable y_k (*e.g.* chemical reactivity) is described in terms of explanatory variables x_{ia} (the 'descriptors' σ, E_s, *etc.*), but no assumption on the relevance of individual variables is required. The method determines the principal components for the descriptors block [equation (2)] and then seeks a simple linear relationship between these components and the property [equation (4)].

$$y_k = \bar{y} + \sum_{a=1}^{A} d_a t_{ak} + e_k \qquad (4)$$

If the descriptor variables were all independent, the number of significant components could equal the number of variables. In this case the numerical solution obtained in PLS would be the same as in MRA. However, in practice, the number of components required is usually much less than the number of variables, owing to the existence of collinearity. Moreover, unlike MRA, PLS is able to detect the existence of subgroups. When the data set shows the presence of subgroups, disjoint PLS models are appropriately used for each group.

Hence, initial PCA enables correct MRA to be carried out, as the assumptions mentioned above can now be justified. However, this two-step procedure (PCA + MRA) can be replaced by a single analysis accomplishing the two steps simultaneously, and this is the basis of the algorithm used in the PLS method (*cf.* refs. 21 and 34).

Principal-component Analysis.—The PCA according to the SIMCA method was carried out on compounds (objects) **a**—**m** in the data matrix reported in Table 5 (unfortunately, insufficient data are available for alkyls **n**—**g** for their use in PCA). A one principal-component (PC) model for the data set describes 70% of the total variance, a second component explaining a further 7%. The PC parameters are listed in Tables 6 and 7.

From Table 7 and Figure 1 (the 'scores' plot), we see that t_1 differentiates the most reactive substituents benzyl, allyl, and methyl and the least reactive group neopentyl from the other primary alkyls and the secondary alkyls which are grouped together. The small influence of t_2 (statistically insignificant according to cross-validation[22]) is apparent from Figure 1.

The b_1 values in Table 6 can be referred to the leaving-group ability of the nucleofuge and confirm the conclusion[2] that

Table 6. Weights, $\bar{x}_i$, b_{i1}, b_{i2}, s_i^2 (1) for reactions (variables) 1—10 in the PCA model

Reactions[a] (variables)	Weights[b]	$\bar{x}_i$[c]	b_{i1}[d]	s_i^2 (1)[e]	b_{i2}[f]
1	1.2701	1.32	0.14	0.64	−0.12
2	0.8568	−0.02	0.23	0.33	0.51
3	0.9121	0.11	0.24	0.14	0.51
4	2.3422	−0.86	0.47	0.13	0.05
5	0.8505	−0.42	0.30	0.06	−0.27
6	1.0567	−0.57	0.46	0.07	−0.17
7	0.4762	−0.64	0.25	0.30	−0.13
8	0.3853	−0.41	0.27	0.07	−0.01
9	0.7289	−0.03	0.22	0.42	−0.57
10	1.3058	−0.27	0.39	0.02	0.15

[a] For definition of reaction 1—10, see Table 5. [b] Factors required to autoscale the log k_2 values for each reaction to the same variance. [c] Arithmetic mean of (relative log k_2) values for relevant reactions. [d] First principal-component loadings for the reactions (variables) 1—10. [e] Variable residual variance, for $A = 1$ see text. [f] Second principal-component loadings for the reactions (variables) 1—10.

Table 7. Principal component scores t_{1k} and t_{2k} for substrates **a**—**m** in the PCA model

Alkyl	*N*-Substituent	t_{1k}	t_{2k}
a	Benzyl	8.85	0.55
b	Allyl	4.53	−0.11
c	Methyl	4.07	−0.77
d	Ethyl	0.98	−0.10
e	n-Propyl	−0.27	−0.38
f	n-Butyl	−0.74	−0.18
g	n-Pentyl	−1.87	−0.56
h	n-Hexyl	−1.20	−0.52
i	n-Heptyl	−1.21	−0.55
j	Neopentyl	−5.53	0.72
k	Isopropyl	−1.65	1.69
l	s-Butyl	−0.74	2.36
m	Isobutyl	−2.77	−0.92

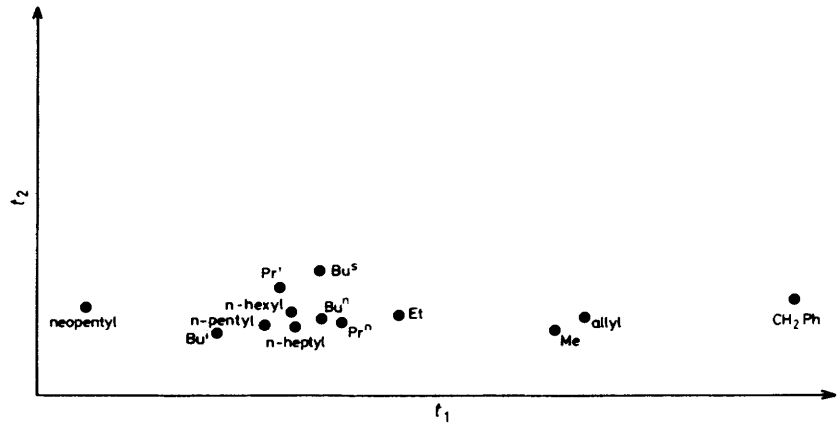


Figure 1. Plot of t_1 *versus* t_2 for compounds (objects) **a**—**m** [PCA model, equation (2)]

Table 8. Descriptors for R^1 and R^2 substituents

Descriptors:[a]	L	B_i	B_{ii}	B_{iii}	B_{iv}	E_s	σ*	MR
H	2.06	1.00	1.00	1.00	1.00	0.00	0.49	1.03
Me	3.00	1.52	1.90	2.04	1.90	−1.24	0.00	5.65
Et	4.11	1.52	1.90	2.97	1.90	−1.31	−0.10	10.3
Pr^n	5.05	1.52	1.90	3.49	1.90	−1.60	−0.12	14.96
Bu^n	6.17	1.52	1.90	4.42	1.90	−1.63	−0.13	19.59
n-Pentyl	7.11	1.52	1.90	4.94	1.90	−1.64	−0.16	24.24
n-Hexyl	8.22	1.52	1.90	5.87	1.90	−1.54	−0.15	28.90
Pr^i	4.11	2.04	3.16	2.76	3.16	−1.71	−0.19	14.96
Bu^s	5.05	1.90	2.76	3.16	3.49	−2.37	−0.21	19.59
Bu^t	4.11	2.59	2.86	2.97	2.86	−2.78	−0.30	19.62
cyclo-C_3H_5	4.14	1.98	2.24	2.88	2.29		0.11	
cyclo-C_6H_{11}	6.17	2.04	3.16	3.49	3.16	−2.03	−0.15	26.69
Ph	6.28	1.70	3.11	1.70	3.11	−3.82	0.60	25.36
Vinyl	4.29	1.60	2.00	1.60	3.90		0.52	10.99
CH=CHMe[b]	5.23	1.90	2.00	1.90	3.09		0.17	15.61

[a] L, B_i, B_{ii}, B_{iii}, B_{iv} taken from ref. 36; E_s, σ*, MR taken from ref. 35. [b] *E* isomer.

Table 9. Relevance of individual descriptors in the PLS models as described by b_i[a] and ψ_i[b]

		Overall 39				Primary 65		Linear 84	
		R^1		R^2		R^1		R^1	
PLS model									
%, V_1[c]	Descriptor	b_i	ψ_i	b_i	ψ_i	b_i	ψ_i	b_i	ψ_i
b_i:	L	−0.26	0.27	0.28	0.33	−0.38	0.19	−0.34	0.98
	B_i	−0.16	0.06	0.29	0.35	−0.22	0.03	−0.35	0.54
	B_{ii}	−0.14	0.04	0.29	0.34	−0.17	0.01	−0.35	0.54
	B_{iii}	−0.28	0.30	0.29	0.34	−0.56	0.58	−0.35	0.55
	B_{iv}	−0.08	0.00	0.29	0.35	0.06	0.00	−0.35	0.54
	E_s	0.12	0.02	−0.29	0.35	0.08	0.00	0.37	0.72
	σ*	0.19	0.12	−0.29	0.36	0.55	0.55	0.37	0.73
	MR	−0.28	0.33	0.29	0.33	−0.39	0.22	−0.34	0.49

[a] Variable loadings; as in Table 6. [b] Modelling power calculated by equation 3. [c] Percentage of the *Y* variance explained by the first PC of the *X* block.

tricyclic (**2**) and pentacyclic (**3**) nitrogen heterocycles are leaving groups as good as chloride and somewhat poorer than bromide ions.

Partial Least squares Analysis.—The relative reactivity in the quinolinium series (**2**) (the most complete one) was chosen as the dependent variable and described as a function of structural parameters ('descriptors') for the alkyls (the *X* block). This analysis is aimed at finding out which descriptors, or what combinations of them, best explain the reactivity data (*i.e.* what effects are responsible for the nucleophilic reactivity of the substrates examined).

As descriptors we used eight parameters available in the literature. For the electronic effect, σ* [35] is clearly appropriate. For steric effects in terms of size, we took the traditional E_s.[35] For the shape, the five Verloop parameters [36] appeared to be the most suitable. The polarizability of each alkyl linked to the carbon atoms undergoing substitution is measured by MR [35] (*cf.* Table 8). To consider simultaneously the primary and secondary alkyls, a second series of eight parameters takes account of the second substituent on the reacting carbon (for the primary substrates this always relates to hydrogen).

The Verloop parameters B_1—B_4 resulting from the STERIMOL computations are usually listed in order of increasing magnitude. However, we have modified this to list them in the order (B_i, B_{ii}, B_{iii}, B_{iv}) of rotation about the *L* (length) axis, commencing with B_i the smallest and with the restrictions that B_{iii} is the opposite to B_i (as indicated in ref. 36) and that $B_{ii} < B_{iv}$ (Table 8). Thus, B_i always corresponds to B_1 as defined by Verloop.

(a) *Overall analysis.* The PLS analysis utilizing the whole data set (17 substrates including 2-methylbutyl and γ-methylallyl) and all 16 descriptors in the *X* block explained only 39% of the variance. This is clearly due to the inhomogeneity of the data set. In Figure 2, where the reactivities (*y*) are plotted against the first component of the *X* block (t_1), two types of deviation from the primary 'normal' alkyls can be observed. The first one

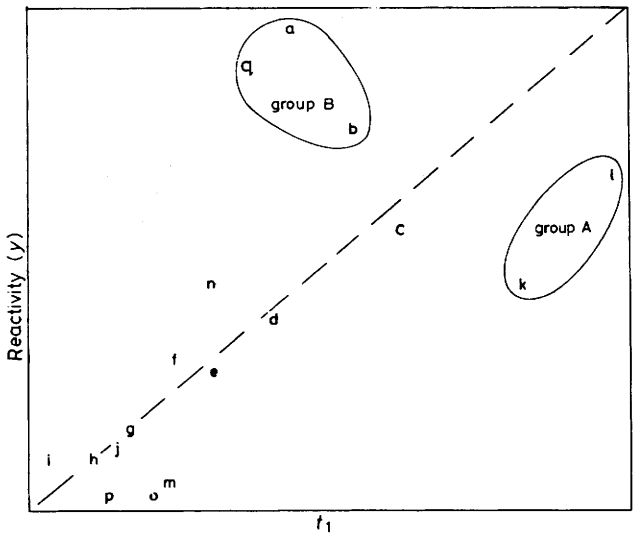


Figure 2. Plot of t_1 *versus y* (reactivity) for compounds (objects) **a**—**q** [overall PLS model, equation (4)]

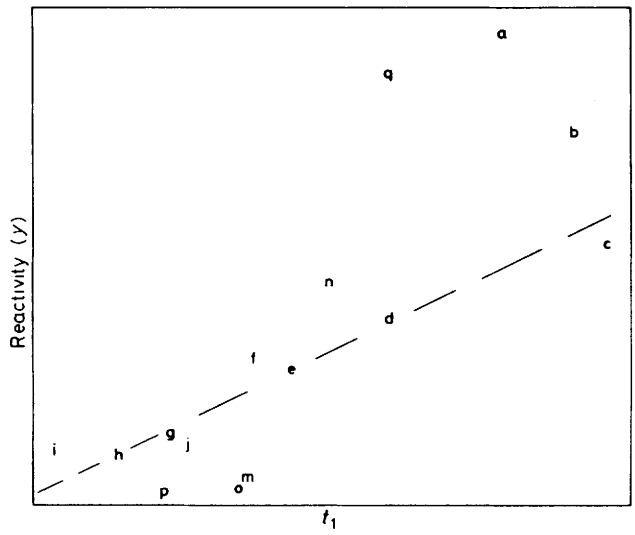


Figure 3. Plot of t_1 *versus* y (reactivity) for compounds (objects) **a**—**j** and **m**—**q** [primary alkyls PLS model, equation (4)]

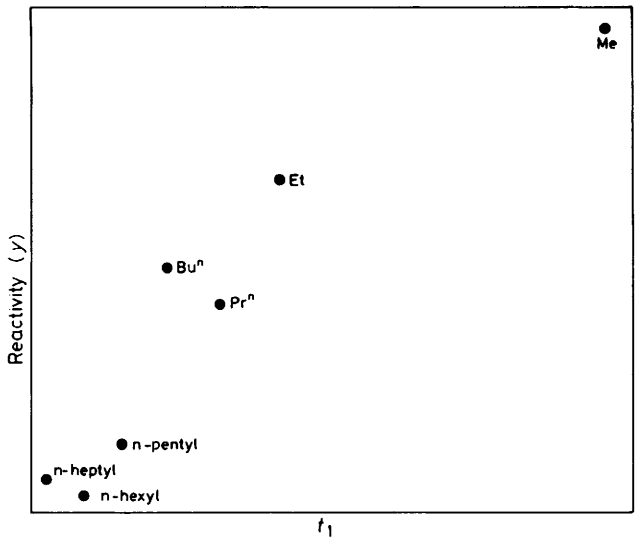


Figure 4. Plot of t_1 *versus* y (reactivity) for compounds (objects) **c**—**i** [linear alkyls PLS model, equation (4)]

involves secondary substrates (group A) and the second one polarizable substituents containing π systems (group B).

Table 9 gives b_i and ψ_i values for the descriptors: the magnitudes for R^2 are greater than for R^1, demonstrating that the presence of a secondary alkyl is that most significant structural modification determining the reactivity of the series. They are all almost equal because of the small variation of the R^2 descriptors throughout the set (*i.e.* only **k** and **l** are different from all the others and very similar to each other). Unfortunately the paucity of the data prevents the application of a disjoint PLS model to this subset.

The most relevant R^1 descriptors appear to be L, B_{iii}, MR, and σ*. This result is also confirmed by the subsequent analysis. Shape is clearly very relevant. The two Verloop parameters found to be most significant represent the substituent length (L) and the dimension (B_{iii}) which is 180° from the smallest one. This may indicate that in the transition state the nucleophile approach is correlated with B_{iii}. Dependence on σ* (the overall size) is not surprising, but that on MR is less easy to interpret.

(b) *Primary alkyls.* A second PLS analysis was carried out excluding the secondary alkyls. Since for all these 15 substrates R^2 = H, eight descriptors define the X block. The results are listed in Table 9 and plotted in Figure 3. The fraction of variance explained goes up to 65%. However, the reactivities of the benzyl, γ-methylallyl, and allyl substrates are higher than predicted by the component for linear primary alkyls **c**—**h** probably because of their polarizability (see the high b value for MR). The branched primary alkyls Bu^i and 2-methylbutyl react slower than predicted.

ψ_i values in Table 9 show that the descriptors relevant to define this first component are again σ*, B_{iii}, L, and MR: the reactivity increases with increasing electronic effect and decreasing steric effect, in agreement with the expected requirements for S_N2 reactions. However, the relevance of two of the Verloop parameters and the small contribution of the bulk steric effect (the E_s parameters) confirms that the steric effect in this series is related to the shape of the substituent rather than to its size.

It is now accepted that steric effects frequently cannot be explained by considering substituents as spheres (*i.e.* by a unique size).[37] Thus, the three-fold symmetry of the methyl group in pyridines was needed to rationalize both the quaternization kinetics and the conformational preference in isopropyl derivatives.[38] Reactivity models for the methylation of substituted pyridines and for the dequaternization of the *N*-methylpyridinium cation have recently been determined[39] and the implications of non-additive steric and electronic effects, as well as the relationships between non-additive kinetics, buttressing effects, and the various steric substituent parameters and models discussed.[40] The results of the PLS analysis provide independent support for the importance of the steric shape of branched primary alkyls in such nucleophilic displacements.

(c) *Linear alkyls.* A further PLS run was carried out considering the linear-chain substrates only (seven compounds). Here also, the model is quite good (84% of variance explained) and descriptors such as σ* and E_s become much more relevant (Table 9). Clearly, in the absence of any branched chain the steric effect is again defined by its size descriptor. The plot of Figure 4 also shows grouping possibly related to the spatial structure of the alkyl chain.

Conclusions.—The new multivariate PCA and PLS methods are a satisfactory alternative to MRA for the study and rationalization of reactivities. This new approach, based on rigorous statistical procedures, enables the empirical treatment of experimental data sets without a built-in bias. The PCA confirms previous findings on the leaving-group ability of quinoliniums (**2**) and acridiniums (**3**) compared with halides and differentiates *N*-alkyl groups capable of resonance delocalization and secondary alkyls from primary alkyls. The PLS analysis points out the importance of the substituent's shape for branched primary alkyls in the nucleophilic reactivity of the quinolinium series (**2**).

Experimental

^{1}H N.m.r. spectra were recorded on a Varian EM360L spectrometer with Me_4Si as internal standard. I.r. spectra were obtained on a Perkin-Elmer 283B spectrophotometer. M.p.s were recorded on a hot-stage apparatus and are uncorrected.

Preparation of Compounds.—5,6-Dihydro-2,4-diphenylnaphtho[1,2-*b*]pyrylium tetrafluoroborate (**4**) (from chalcone, α-tetralone, and boron trifluoride–ether[6]) had m.p. 270 °C (lit.,[6] 270 °C); 5,6,8,9-tetrahydro-7-phenyldibenzo[*c*,*h*]xanthylium tetrafluoroborate (**5**) (from 2-benzylidene-α-tetralone, α-tetra-

lone, and boron trifluoride–ether[6]) had m.p. 258—260 °C (lit.,[6] 265 °C).

General Procedure for Preparation of Pyridinium Salts (*Table* 1).—In a typical experiment, amine (2.4 mmol) was added dropwise to a suspension of pyrylium tetrafluoroborate (1.2 mmol) in CH_2Cl_2 (15 ml) and the deep red mixture was stirred at 25 °C for the time given. The colour changed to dark green. Solvent (8—10 ml) was removed (50 °C; 20 mmHg) and the residue was treated with ether (50 ml) to give the product [for compound (**20**), exceptionally n-hexane was used in place of ether]. Washing the product with warm water and then diethyl ether removed amine salts. The dried residue (60 °C and 1 mmHg) was dissolved in Me_2CO and reprecipitated with ether. Physical and spectroscopic properties are recorded in Tables 1 and 2.

Kinetic Measurements.—The kinetics were followed by u.v. spectrophotometry under psuedo-first-order conditions using the procedure already described.[9] The concentration of quinolinium or acridinium was 6.4×10^{-5} mol l^{-1}, while those of piperidine ranged from 0.0008 to 0.32 mol l^{-1}. Pseudo-first-order rate constants were calculated from the slope of the plot of ln (D_o/D) at the wavelengths reported in Table 3 *versus* time. Second-order rate constants were calculated from the slope and first-order rate constants from the intercept of the plot of k_{obs} *versus* piperidine concentration. For the definition and calculation of errors and for the estimation of the precision of k_{obs}, see ref. 41.

Acknowledgements

We thank Professor S. Wold, Umea University, for providing the SIMCA package, the Commission for Educational Exchange between the United States and Italy for a Fulbright travel grant (to G. M.) and Ministero della Pubblica Istruzione for financial support (to S. C. and G. M.).

References

1 Part 17, A. R. Katritzky, Z. Dega-Szafran, M. L. Lopez-Rodriguez, and R. W. King, *J. Am. Chem. Soc.*, 1984, **106**, 5577; A. R. Katritzky, K. Sakizadeh, B. Gabrielsen, and W. le Noble, *J. Am. Chem. Soc.*, 1984, **106**, 1879 is considered as Part 18.
and R. W. King, *J. Am. Chem. Soc.*, 1984, **106**, 5577.
2 A. R. Katritzky, K. Sakizadeh, Y. X. Ou, B. Jovanovic, G. Musumarra, F. P. Ballistreri, and R. Crupi, *J. Chem. Soc., Perkin Trans. 2*, 1983, 1427.
3 A. R. Katritzky, Y. X. Ou, and G. Musumarra, *J. Chem. Soc., Perkin Trans. 2*, 1983, 1449.
4 C. K. Ingold, 'Structure and Mechanism in Organic Chemistry,' Bell, London, 1969, 2nd edn., p. 555.
5 A. Streitwieser, 'Solvolytic Displacement Reactions,' McGraw-Hill, New York, 1962, p. 13.
6 A. R. Katritzky and S. S. Thind, *J. Chem. Soc., Perkin Trans. 1*, 1980, 1895.
7 For a review see A. R. Katritzky, *Tetrahedron*, 1980, **36**, 679.
8 A. R. Katritzky and S. S. Thind, *J. Chem. Soc., Perkin Trans. 1*, 1980, 865.
9 A. R. Katritzky, G. Musumarra, K. Sakizadeh, and M. Misic-Vukovic, *J. Org. Chem.*, 1981, **46**, 3820.
10 A. R. Katritzky, A. M. El-Mowafy, G. Musumarra, K. Sakizadeh, C. Sana Ullah, S. M. M. El Shafie, and S. S. Thind, *J. Org. Chem.*, 1981, **46**, 3823.
11 A. L. Solomon and H. C. Thomas, *J. Am. Chem. Soc.*, 1950, **72**, 2028.
12 G. Le Ny, G. Roussi, and H. Felkin, *Bull Soc. Chim. Fr.*, 1965, 1920.
13 E. E. Royals and A. H. Neal, *J. Org. Chem.*, 1956, **21**, 1448.
14 B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, 1972, **94**, 5632.
15 C. Albano, W. J. Dunn, U. Edlund, E. Johansson, B. Norden, M. Sjöström, and S. Wold, *Anal. Chim. Acta*, 1978, **103**, 429.
16 D. L. Massart, A. Kijkstra, and L. Kaufman, 'Evaluation and Optimization of Laboratory Methods and Analytical Procedures,' Elsevier, Amsterdam, 1978.
17 N. Draper and M. Smith, 'Applied Regression Analysis,' Wiley-Interscience, New York, 1981, 2nd. edn.
18 K. V. Mardia, J. T. Kent, and J. M. Biddy, 'Multivariate Analysis,' Academic Press, London, 1979.
19 S. Wold and M. Sjöström, 'SIMCA: a Method for Analyzing Chemical Data in Terms of Similarity and Analogy,' in B. R. Kowalski, 'Chemometrics: Theory and Application,' ASC Symp. Series No. 52, American Chemical Society, Washington, 1977, p. 243.
20 C. Albano, G. Blomquist, D. Coomans, W. J. Dunn, U. Edlund, B. Eliasson, S. Hellberg, E. Johansson, B. Norden, M. Sjöström, B. Soderstrom, H. Wold, and S. Wold, 'Pattern Recognition by Means of Disjoint Principal Components Models: Philosophy and Methods,' in 'Symposium i Anvent Statistik,' eds. A. Hoskuldsen and K. Esbensen, Copenhagen, 1981, p. 183.
21 S. Wold, C. Albano, W. J. Dunn, K. Esbensen, S. Hellberg, E. Johansson, and M. Sjöström, 'Pattern Recognition: Finding and Using Regularities in Multivariate Data,' in 'Food Research and Data Analysis,' eds. H. Martens and H. Russwurm, Applied Science, London, 1983, p. 147.
22 S. Wold, *Technometrics*, 1978, **20**, 397.
23 M. Sjöström and S. Wold, *Acta Chem. Scand., Ser. B*, 1981, **35**, 537.
24 (*a*) C. Albano and S. Wold, *J. Chem. Soc., Perkin Trans. 2*, 1980, 1447; (*b*) S. Alunni, S. Clementi, U. Edlund, D. Johnels, S. Hellberg, M. Sjöström, and S. Wold, *Acta Chem. Scand., Ser B*, 1983, **37**, 47.
25 D. Johnels, S. Clementi, W. J. Dunn, U. Edlund, H. Grahn, S. Hellberg, M. Sjöström, and S. Wold *J. Chem. Soc., Perkin Trans. 2*, 1983, 863.
26 U. Edlund and A. Nostrom, *Org. Magn. Reson.*, 1977, **9**, 196.
27 U. Edlund and S. Wold, *J. Magn. Reson.*, 1980, **37**, 183.
28 G. Musumarra, S. Wold, and S. Gronowitz, *Org. Magn. Reson.*, 1981, **17**, 118.
29 G. Musumarra, G. Scarlata, and S. Wold, *Gazz. Chim. Ital.*, 1981, **111**, 499.
30 M. Sjöström and S. Wold, *J. Chem. Soc., Perkin Trans. 2*, 1981, 104.
31 G. Musumarra, F. P. Ballistreri, S. Muratore, A. R. Katritzky, and S. Wold, *J. Chem. Soc., Perkin Trans. 2*, 1982, 1049.
32 W. J. Dunn and S. Wold, *Bioorg. Chem.*, 1980, **9**, 505.
33 H. Wold, 'Systems under Indirect Observation, Causality–Structure–Prediction,' eds. K. G. Joreskog and H. Wold, North Holland, Amsterdam, 1981, vol. 2.
34 S. Wold, H. Wold, W. J. Dunn, and A. Ruhe, 'The Collinearity Problem in Linear and Non-linear Regression. The Partial Least Squares Approach to Generalized Inverses,' Report UMINF 83.80, ISSN 0348-0542, Umea University, Institute of Information Processing and Chemistry, 1980.
35 C. Hansch and A. Leo, 'Substituent Constants for Correlation Analysis in Chemistry and Biology,' Wiley-Interscience, New York, 1979.
36 A. Verloop, W. Hogenstraaten, and J. Tipker, *Drug Design*, 1976, **7**, 165.
37 For a recent review see: R. Gallo, *Prog. Phys. Org. Chem.*, 1983, **14**, 115 and references therein.
38 C. Roussel, A. T. Balaban, V. Berg, M. Chanon, R. Gallo, G. Klatte, J. A. Memiaghe, J. Metzger, D. Oniciu, and J. Pierrot-Sanders, *Tetrahedron*, 1983, **39**, 4209.
39 J. C. Schug, J. W. Viers, and J. I. Seeman, *J. Org. Chem.*, 1983, **48**, 4892.
40 J. I. Seeman, J. W. Viers, J. C. Schug, and M. D. Stovall, *J. Am. Chem. Soc.*, 1984, **106**, 143.
41 A. R. Katritzky, Y. X. Ou, J. Ellison, and G. Musumarra, *J. Chem. Soc., Perkin Trans. 2*, 1983, 1421.

Received 17*th September* 1984; *Paper* 4/1600